AL-TP-1991-0003

AD-A240 808

ARMSTRONG LABORATORY

SURVIVAL ANALYSIS: A TRAINING DECISION APPLICATION

Julia A. Stephenson

Department of Business Computer Information Systems
and Management Science
College of Business
University of North Texas
Denton, TX 76203

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

August 1991

Interim Technical Paper for Period June 1990 - March 1991

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

NOTICES

This paper has been reviewed and is approved for publication.

J. SCOTT NEWCOMB
Project Scientist

HENDRICK W. RUCK, Technical Director
Technical Training Research Division

RODGER D. BALLENTINE., Colonel, USAF
Chief, Technical Training Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
August 1991

3. REPORT TYPE AND DATES COVERED
Interim Paper – June 1990 – March 1991

4. TITLE AND SUBTITLE
Survival Analysis: A Training Decision Application

5. FUNDING NUMBERS
PE - 65502F
PR - 1121
TA - 10
WU - 66

6. AUTHOR(S)
Julia A. Stephenson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Business Computer Information Systems and Management Science
College of Business
University of North Texas
Denton, TX 76203

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Armstrong Laboratory
Human Resources Directorate
Technical Training Research Division
Brooks Air Force Base, TX 78235-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL-TP-1991-0003

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Scott Newcomb, (512) 536-2242

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The life of a task in an airman's inventory of tasks performed has not been investigated. How long a task remains (survives) in an individual's task inventory is of interest for training purposes. Survival analysis can possibly be used to measure task survivability. However, survival analysis uses longitudinal data whereas the USAF Occupational Survey Program captures vertical data (i.e., a snapshot is taken of the work force at one moment in time). Nonetheless, because survival analysis can incorporate both time and censored data, it could provide useful information about task survivability. In this effort, a task survival data base was modeled by combining both occupational survey data and known attribution data. Survival analysis functions were then generated. Results show both that survival analysis can be used to study task survivability and that this approach produces accurate estimates of task life.

14. SUBJECT TERMS
survival analysis
task cycle
task life
task perishability
task survivability job task analysis

15. NUMBER OF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

This paper summarizes the initial investigation into the use of survival analysis with occupational survey data. This research was conducted under the United States Air Force Summer Faculty/Graduate Student Research Program and was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under contract F49620-88-C-0053.

The author would like to thank the Human Resources Directorate, and in particular the Technical Training Research Division, for providing an environment supportive of this work. Several individuals were of specific help: Colonel Rodger Ballentine; Drs. Scott Newcomb, Hendrick Ruck, and Wesley Regian; Capt Mary Knoll; Mr. Winston Bennett; and Ms. Melody Darby. To the rest of the Division, the Directorate Library staff, and the Information Sciences Division, thanks for all of your assistance. Finally, a special thanks to Dr. Stanley D. Stephenson from the School of Business at Southwest Texas State University for both his help with the occupational survey portion of this paper and his assistance throughout this entire project.

SURVIVAL ANALYSIS: A TRAINING DECISION APPLICATION

SUMMARY

The life of a task in an airman's inventory of tasks performed has not been investigated. Yet, knowledge of how long a task remains (survives) in an individual's task inventory is of interest, primarily for training purposes. Survival analysis, an analytical technique frequently used in the biomedical field, could possibly be used to measure task survivability. However, survival analysis uses longitudinal data whereas the USAF Occupational Survey Program captures vertical data (i.e., a snapshot is taken of the work force at one moment in time). Nonetheless, because survival analysis can incorporate both time and censored (incomplete) data, it could provide useful information about task survivability. In this investigation, a task survival data base was modeled by combining both occupational survey data and known attrition data. Survival analysis functions were then generated. Results show both that survival analysis can be used to study task survivability and that this approach produces accurate estimates of task life. Theoretical implications and further applications are discussed.

I. INTRODUCTION

A general goal of organizations is to efficiently manage the productivity of its workers. Thus, knowledge of the cycle of a task in a person's job inventory, ir particular task emergence and perishment, would seem to be of value to any organization. However, little research has been conducted on task life. This lack of research could be attributed to the method in which job inventory data are collected. Or, perhaps, the time dimension involved in a task cycle has simply not been viewed as informative or necessary for understanding what actually occurs in a job or career field.

Knowledge about the task cycle would be extremely useful to an organization for several reasons. Training decisions could be more effectively implemented. Knowing the average life expectancy of a task in a job inventory at a particular point in time may give insight into cross-training decisions.

Also, improved decisions on where to train, formal schooling or on-the-job (OJT), could be made with a better understanding of the probabilities associated with performance of particular tasks. If a low percentage of job incumbents are performing a task, perhaps formal training on that task for all workers in that field is unnecessary. OJT for only those job incumbents who need the information would be more cost effective.

Another area of occupational research which is directly tied to the task cycle is skill decay. Two functions that would help predict skill decay are task emergence and task perishment. Workers experience skill decay if they no longer hold the task in their inventory. Also, if they are trained on a particular task just before beginning the job but do not start performing that task until they have been working for a year, the skills necessary for that task will decay.

This paper reports on the feasibility of analyzing the task cycle, in particular task perishability, using the statistical technique of survival analysis. The application of this study is for the type of task typically found in an Air Force occupational survey. Thus, an overview of the Air Force Occupational Survey Program is given first. This section is followed by an introduction to survival analysis and then by the methodology of this study in applying survival analysis to Air Force occupational data. Results are then presented, and the paper concludes with a discussion of implications and recommendations for further research.

II. USAF OCCUPATIONAL SURVEY PROGRAM

There are many forms and types of job analysis. One of the most widely accepted and used methods is job-task inventory analysis (Levine, Ash, Hall, & Sistrunk, 1983). This method involves developing a task list containing every task that workers in the particular work specialty could possibly perform. The job-task inventory is then administered in survey form to workers, and the workers answer two questions about each task: Do you perform this task, and, if so, how much time do you spend performing this task compared with the other tasks you perform? Percent members performing and relative time spent on task are then combined with other information, such as task difficulty, task criticality, and knowledge from subject-matter experts (SMEs) to determine the training emphasis for each task. As might be imagined, a major use of the results of the job-task method has been in making training decisions. Training programs have been either reduced or expanded based on whether or not tasks are being performed on the job.

The United States Air Force (USAF) developed this job-task inventory analysis over a 20-year period. To analyze the data obtained, the USAF developed a series of computer programs called the Comprehensive Data Analysis Programs (CODAP), and the USAF job analysis program is now frequently referred to by the term, CODAP (or TI/CODAP) (Christal & Weissmuller, 1988).

The USAF divides career fields into Air Force Specialty Codes (AFSCs); e.g., all jet engine mechanics are grouped together. CODAP usually involves taking a snapshot of an entire ASFC at one point in time. Therefore, the data are vertical rather than longitudinal. Consequently, the data do not provide information about what an individual worker does over a 20-year career. Instead, CODAP provides information about what all workers are doing in specific time intervals. Typically, an Occupational Survey Report (OSR) will provide task performed information broken down by term of enlistment. Airmen in their first 48 months are first-term enlistees, and airmen in months 48 to 96 months are second-term enlistees. Those members who have been in the Air Force for longer than 96 months are considered career enlistees.

A second important component of the occupational survey program is that, for AFSCs with less than 3,000 members, the Air Force administers the job-task inventory to 100 percent of the work force. This method of surveying produces a response rate of 80%, which is basically all job incumbents available for work when the survey is administered.

The method of surveying and the high rate of return for the inventories are significant for one important reason. In essence, the USAF job analysis program produces population parameter information about the percent of workers who are performing a task at a particular point in time. Thus, if the data show, for example, that 40% of the workers at the 4-year (48-month) point are performing a task, then the 40% figure can be treated as a parameter as opposed to an estimate.

III. SURVIVAL ANALYSIS

A. Background

Although survival analysis is perhaps new to behavioral scientists, it has a history of use in other disciplines, primarily the biomedical field where it has been used to study the effectiveness of a treatment (e.g., survival of cancer patients after treatment). In electrical engineering, survival analysis is labeled as a reliability study and is used to measure the failure rates of electrical components. Because survival analysis has such strong roots in biostatistics and engineering, it is not surprising that most of the survival analysis textbooks are slanted toward handling either medical or electrical component data.

A major contribution in the history of survival analysis was the standarization of the method of estimating survival probability; this was accomplished with the Kaplan and Meier (1958) product limit estimator. A second major reference is the article by Cox (1972) in which survival analysis was extended to include a regression component.

During the 1980s, this type of analysis has been used extensively in other fields. As is often the case in the diffusion of knowledge from the theoretical and quantitative sciences to the applied sciences, the process has been slow. As survival analysis has spread to other disciplines, it has taken on new names. Economists use survival analysis to conduct duration studies (Kiefer, 1988; Ridder, 1990), and sociologists have coined the term event history analysis (Tuma, Hannan, & Groeneveld, 1979). This technique has also appeared in business journals (e.g., analyzing employee turnover) (Darden, Hampton, & Boatwright, 1987; Morita, Lee, & Mowday, 1989).

An important point of this discussion is that survival analysis has a well-documented history. Moreover, many popular statistical software packages contain survival analysis modules (Goldstein et al., 1989).

Survival analysis enables the researcher to determine probabilities associated with the length of time for a binary, dependent variable to change states. The only required independent variable is the time that expires from the start of the experiment to the change in state of the dependent variable. Both the origin time and the exact point at which the dependent variable changes must be precisely defined (Cox & Oakes, 1985). Two other assumptions are that the sample is homogeneous and that the length of time for the dependent variable to change states is a positive value (Lawless, 1982).

One of the strengths of survival analysis is the ability to include some information about the censored data. Censoring may occur if the experiment ends before the dependent variable changes. For instance, a medical follow-up study may be funded for only 5 years. Those patients who are still alive when the study ends are censored because their actual time of death is not known. Another type of censoring occurs when an item leaves the sample before termination of the experiment without the dependent variable having changed states (e.g., a patient moves out of state and is no longer part of the study). In most parametric statistical analyses, these data would have to be omitted from the sample. However, the fact that the item had not changed at the point of leaving or ending the experiment does provide some relevant information that should be incorporated into probabilities associated with the time at which the dependent variable changes states.

Two assumptions of censoring are inherent in survival analysis: (a) The censoring of one item does not affect censoring or the length of time to the change of state in the dependent variable in any other items; and (b) censoring times are non-informative (Lagakos, 1979). Censoring is informative if the item leaves the sample for reasons directly tied to the experiment. An example of informative censoring is a medical patient removing himself because of side effects of the treatment.

B. Theory

Survival analysis incorporates several related statistics. For the purposes of this paper, T will represent the minimum among (a) the length of time between the start of the "experiment" and the change in state of the dependent variable, (b) the length of time between the start of the experiment and the time at which an item leaves the sample without experiencing a change in the dependent variable, and (c) the length of time between the start and end of the experiment. The probability density function is the probability that the dependent variable changes at time t:

$$f(x) = P(T=t). \tag{1}$$

The failure and survival functions represent cumulative distributions of the change in the dependent variable. Failure is defined as the change in the dependent variable; survival is the lack of change:

$$F(t) = P(T<t) = \int_0^t f(x)dx, \text{ and} \tag{2}$$

$$S(t) = P(T>t) = 1 - F(t) = \int_t^\infty f(x)dx. \tag{3}$$

The hazard function represents the conditional probability that the dependent variable will change in time t, given that it had not changed prior to time t:

$$h(t) = P(T=t \mid T>t) = \frac{f(x)}{S(t)} = \frac{-S(t)}{S(t)}. \tag{4}$$

The cumulative hazard function, used primarily for comparison among samples, is the integral of the hazard function.

$$H(t) = h(x) = \frac{-S'(x)}{S(x)} = -\log[S(t)]. \tag{5}$$

The mean life residual represents the average length that the dependent variable will survive beyond time t (Oakes & Dasu, 1990):

$$r(t) = E[T-t \mid T>t] = \frac{\int_t^\infty S(x)}{S(t)}. \tag{6}$$

C. Estimation

Beyond the theoretical relationships among these probabilities, there are two methods of estimating the survival curve;the Kaplan-Meier (1958) method and the life-table method (Lawless, 1982). The primary difference between these two estimators is how they handle the censored data.

Under the Kaplan-Meier method, the hazard function, the probability of an individual component of the sample dying at time t, given survival up to point t, is estimated as

$$h(t) = \frac{d_t}{n_t}. \tag{7}$$

where n_t is the number at risk (i.e., the number in the sample who have yet to change states in the binary variable) at time t and dt is the number of observations in the sample whose binary, dependent variable has changed states at time t. Thus, the probability of surviving at time t, given survival up to point t, will be

$$p(t) = 1 - h(t). \tag{8}$$

The survival function is estimated using simple Bayesian probabilities (i.e., P(A and B) = P(B A) P(A)). For example, by substituting the event of surviving t=1 as A and surviving t=2 as B, the probability of surviving A and B will be the conditional probability of surviving t=2, given survival at t=1, multiplied by the probability of surviving t=1. This procedure can be extended to the end of the time frame, thus estimating the entire survival function. In mathematical terms, the survival function is

$$S(t) = \prod p(t). \tag{9}$$

This method includes all of the censored data in the number at risk at each time period. If precise measurement of the time of change of the dependent variable is possible, d_t will be one. Because the data are continuous, the event of two changes in the dependent variable at exactly the same point in time is not possible.

The life-table, or actuarial, method is the second type of estimator for the survival function. The hazard function is estimated as

$$h(t) = \frac{d_t}{n'_t}. \tag{10}$$

where n'_t represents the number of people at risk minus one-half of the censored data at time t. The complement of h(t), the probability that a person survives time t given that he survived time t-1, and the estimator for the survival function are the same as the Kaplan-Meier method.

The difference in these two methods is evident. With the Kaplan-Meier method, all of the censored data points remain in the risk set during the time period at which they leave the sample. In the life-table method, half of the censored data points are removed from the risk set. The reason for this difference lies in the type of data under analysis. The Kaplan-Meier estimator is the standard method for treating continuous data, whereas the life table method is typically used for discrete data. Because discrete data are usually analyzed in interval form, removing some of the censored data points from the risk set at each interval is logical, especially as the width of the interval increases. Depending on the variable being examined, one may have reason to believe that all of the censored data points would not survive beyond the bounds of the interval. For instance, a typical use of the interval method involves determining risks and probabilities of occurrence for intervals of at least one full year. The likelihood of an item censored at the beginning of an year surviving through the whole year may be low. On the other hand, items which are censored at the middle or end of the year would have a higher probability of surviving the year.

IV. METHOD

At a first glance, survival analysis does not seem appropriate for examining Air Force occupational data. The actual time when a person stops performing a specific task is not recorded. Another problem is that few data are maintained on persons who leave the service. Upon closer examination, however, data gathered by the occupational surveys do meet the required survival analysis assumptions.

The binary, dependent variable is the task's either being in an incumbent's inventory or not being in an incumbent's inventory. In the occupational survey, the person checks if a specific task is currently being performed. For all tasks checked, the person then indicates the relative amount of time spent on each task. The potential problem of an incumbent's not marking tasks that are only occasionally performed is overcome by providing the job incumbent with the opportunity to indicate relative time spent.

The second assumption of survival analysis is that the origin and exact point at which a task leaves a person's inventory must be specified. Actually, however, the only necessary requirement is knowing the length of time that the task is in the job inventory. To meet this requirement, a small mental transformation of the occupational survey data is necessary. The occupational survey gives for each time interval the number and the percentage of people who currently hold the task in their inventory. The difference between two intervals in the number/percentage of people who do not hold the task in their inventories is, in effect, an indication of the number of deaths (those who have changed the states of the dependent variable) during that interval. In referring back to the product limit estimator (equation 7), this number becomes the d value (i.e., the number of deaths).

Therefore, occupational surveys meet the primary assumptions of survival analysis. The problem area is the inclusion of the censored data. Although the Air Force does have information regarding attrition rates, whether the specific task is in the person's inventory when leaving the service is unknown.

For the purposes of this study, I generated a 1,000-person data base. This data base included actual data points for a task leaving an airman's job inventory, as well as censored data which simulated those airmen who leave the Air Force prior to the task leaving their inventory. In this model, all of the airmen had either stopped doing the task or had left the Air Force by the 72th month. Finally, of the 1,000-airman data base, 300 (30%) were considered censored (i.e., these airmen left the service or career field before the task dropped from their task inventory).

Though this model is not specific to any one career field, it does incorporate several facts that are intrinsic to job/career development in the Air Force. For instance, airmen often undergo basic military training and formal schooling during the first 12 months of their military career. Thus, the model starts at the 13th month, which is actually the first point in time that a task could leave an incumbent's inventory.

Another consideration is the large change in status at the 48th month. At this point, many airmen (up to 50%) decide not to re-enlist. Of those who do continue in the Air Force, some change career fields. These changes result in many censored data points at the 48th month. The model accounts for this large change by placing 200 of the total 300 censored data points at the end of the first enlistment. This figure does not mean that only 20% of the airmen left the service at the 48th month but that 200 left this career field and still held the task in their inventories. In reality, some of the airmen who leave the service/career field at the 48th month had already stopped performing the task. Thus, they would have been included as observed (non-censored) changes in earlier months. The remaining 100 censors were randomly distributed throughout the 72 months, with a higher probability of occurrence prior to the 48th month versus after the 48th month.

This type of information is obviously discrete, interval data. As discussed earlier, the only difference in the discrete and the continuous estimators of the survival function is in the handling of the risk set. The continuous method includes all of the censors for a particular point in time in the risk set. The discrete method removes half of the censors at each point from the risk set for that point in time. However, the interval length in a typical survival analysis study using discrete data is 1 year. Therefore, the 1-month interval used in this study, when compared to the typical life table interval of a full year, more closely approximates continuous data. Because of the intrinsic qualities of the data and because removing half of the censors from the risk set did not seem appropriate for this study's time interval, the more commonly used Kaplan-Meier estimator for continuous data was used to estimate the survival function.

The survival function was generated through the use of Proc Lifetest in Statistical Analysis System (SAS). The hazard and mean life residual functions were calculated using the results of Proc Lifetest and various data steps and basic procedures, also in SAS (see the Appendix).

V. RESULTS

Figure 1 shows the survival function for the model database. It represents the probability of an airman performing the task at a specific time period. For example, at the 36th month, the probability that an airman will still be performing this task is 0.54.

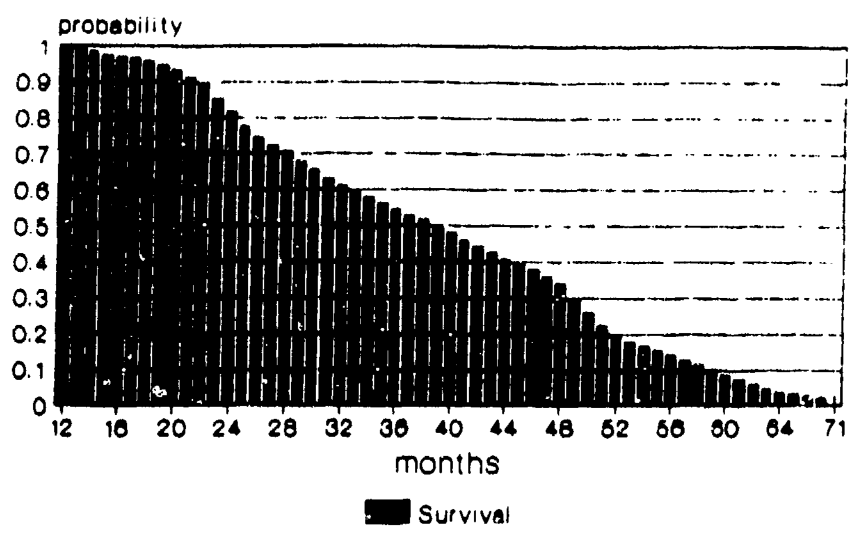

Figure 1. **Survival Function.**

Figure 2 represents the mean life residual function for the data base. This function can be interpreted as the average length that an airman will be performing the task beyond a specific time period.

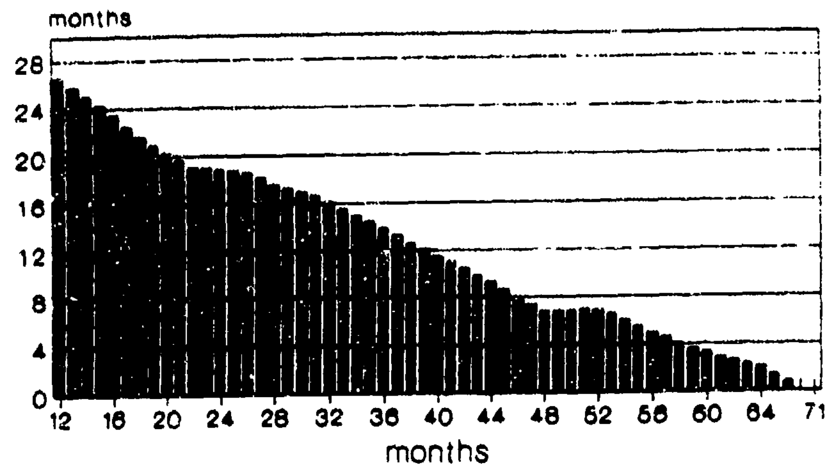

Figure 2. **Mean Life Residual.**

At the 36th month, an airman will be performing this task an average of 13.8 more months. The large amount of naturally occurring censored data at the 48th month severely affects this function. Intuitively, one may wonder how the function remains level immediately following the 48th month. This result is linked directly to the dramatic decrease in the survival function during this time period, a decrease due to the large amount of censoring.

The hazard function, which is the probability that an airman will stop performing the task in a specific time period, given that the task was in his inventory in the preceding month, is exhibited in Figure 3. In this model, the probabilitv that the task will leave an airman's inventory in month 36 is 0.028. At month 71, the last person in the data base stops performing the task; thus, the hazard rate is 1.00. Although, the hazard function begins to rise after the 48th month because of the decrease in the risk set, this function is not necessarily autocorrelated.

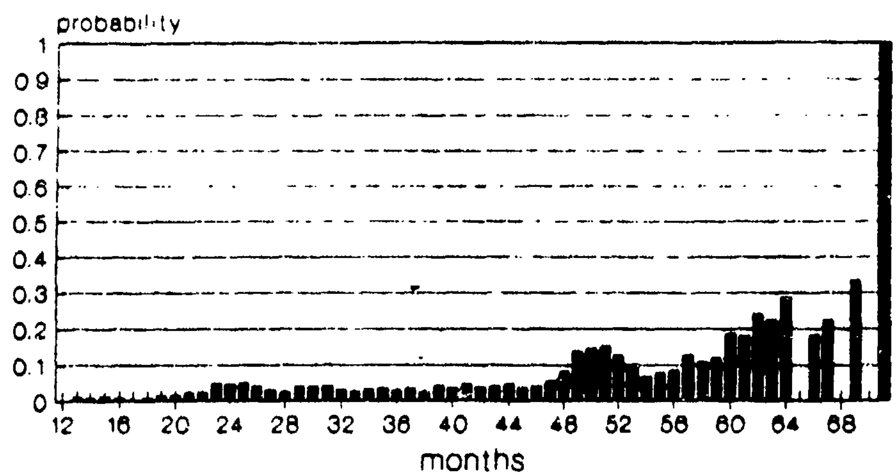

Figure 3. **Hazard Function.**

Figures 4 and 5 show a comparison between the data base with all 1,000 airmen (survival analysis) and the data base with 700 airmen (censored data omitted/conventional analysis). The difference in the two survival functions (Figure 4) is greatest at the 48th month, the point of the heaviest censoring. At this point in time, survival analysis is simply working with more information (i.e., it included partial information from the censored data) and can provide a more accurate estimate of the survival function. After the 48th month, as the size of the two data bases converge, the two curves become more similar.

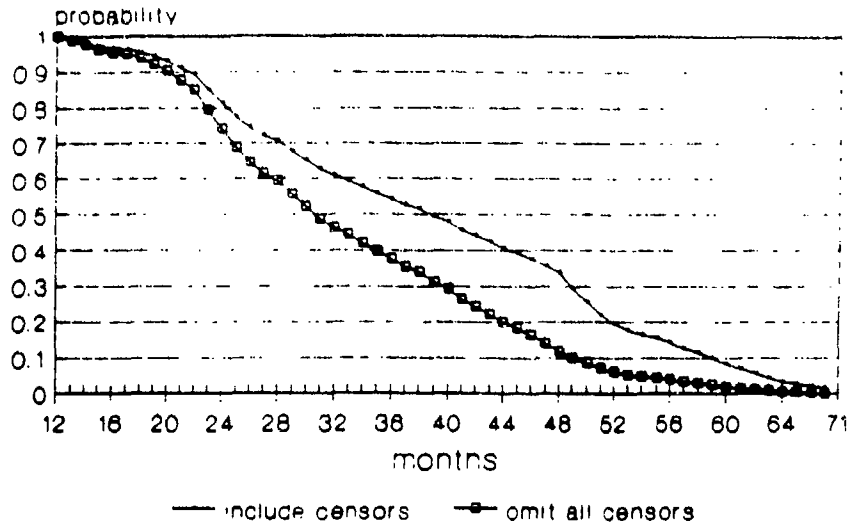

Figure 4. **Survival Comparison.**

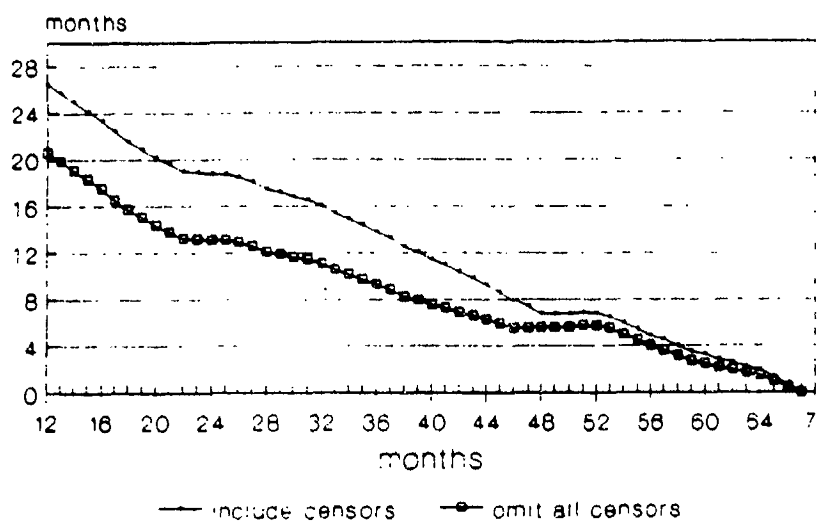

Figure 5. **Mean Life Comparison.**

The difference between the two mean life residual functions (Figure 5) is greatest at the beginning of the 13th month. This result is observed because removing the 300 censored data points at the start of the study decreases the risk set and thus lowers the probability of survival. Recall from Equation 6 that this function is the sum of the survival function from time t to the end of the study, divided by the survival function at time t. At the 48th month, both samples have almost the same number of people, which causes the two curves to become very similar. Thus, censoring after the first term of enlistment has less effect on the mean life residual function.

The data from Figures 4 and 5 could also be presented in a table format, as shown in Table 1. This table gives an example of the survival and the mean life residual functions for months 36 through 40.

Table 1. **Comparison Data**

	Survival Function		**Mean Life Residual**	
Month	**Include Censors**	**Omit Censors**	**Include Censors**	**Omit Censors**
36	.544	.377	13.807	9.273
37	.527	.354	13.264	8.871
38	.516	.340	12.547	8.244
39	.495	.313	12.087	7.959
40	.479	.293	11.472	7.502

VI. DISCUSSION

The results of this study show that survival analysis can be used to investigate task perishability. Due to the method of collecting task data in the Air Force Occupational Survey Program, accurate figures can be obtained for the change in state of the binary variable (i.e., task performed). Historical attrition data are available for all career fields. Thus, censoring is the only unknown variable, and it can be accurately estimated by combining occupational and attrition data. Therefore, an appropriate data base can be created for any AFSC.

The results of the analysis also show the advantage of using survival analysis to measure task perishability. Figures 4 and 5 vividly illustrate the difference in analyzing task perishability using survival analysis, which can accommodate censored data, and using conventional analytical procedures, which essentially discard censored data. Estimations of both the survival and mean residual life functions are more accurate using survival analysis. Therefore, the results of this study strongly suggest that analyzing task perishability with survival analysis should continue to be studied.

The use of survival analysis to examine occupational data, such as task perishability, is a new application of this statistic. Thus, several research issues need further examination. Of primary concern is the inclusion of censored data. This use of survival analysis meets the model assumption that the censoring is non-informative (i.e., airmen do not leave the service because of a specific task). However, as mentioned earlier, the Air Force does not maintain records of the tasks performed by persons who do leave. Therefore, determining the number of censored data points at each interval will always have to be modeled. Also, the typical survival analysis study does not include a period of concentrated censoring as is the case with Air Force data. Nonetheless, the logical start point still would be to use the known information on percent (of those who complete the occupational survey) members performing as an estimation of the percentage of those who have left the career field but still hold the task in their inventories. However, various censoring models should be analyzed.

Survival analysis assumes that 100% of the airmen are performing the task from the start. If 100% are not performing the task at the start of the study, it raises a potential theoretical issue. However, the math underlying the model is primarily based on conditional probabilities; thus, deviating from this assumption may not have a severe effect on the task performance probabilities. Nonetheless, this issue should be studied.

Another theoretical question concerns the homogeneity of the airmen in a particular career field. Occupational data are vertical rather than longitudinal; thus, this assumption is important in making inferences on the results of the analysis. One is wary of assuming that airmen in their 20th year are homogeneous with first-term enlistees. Because of many variables, the Air Force is very different now than it was 20 years ago. However, a snapshot view of today's work force may give a more homogeneous sample of tasks currently being performed than would a longitudinal study. Information on the current status of a task in a career field is more relevant to making sound decisions for the future than is information on the task from an earlier point in time.

The previous issues have been theoretical in nature. There are also more applied issues. For instance, a more accurate analysis of when a task leaves an airman's job inventory may be accomplished by subgrouping the career field with a covariate such as present grade, skill level, or gender. Another significant covariate may be first versus second career field. One would presume that, because of differences in present grade, many low-level tasks will be performed longer by enlistees in their first career field than those in their second. Other covariates might be percent members performing and task difficulty.

Another area of interest for further research is task emergence. The model set forward in this study could easily be restructured to analyze when a task enters a job inventory (i.e., the change in the dependent variable would be from not performing a task to performing the task). This approach would negate the earlier mentioned "100% performing at the start" theoretical issue. Also, a more realistic situation in the USAF is that none of the airmen are performing the task on the first day of the job.

Understanding when a task begins to emerge in a particular career field also has several training implications. For instance, if a task does not emerge until after a year on the job, perhaps the task should not be trained until that point. The cost of training during the formal school (just prior to the airman's entering the career field) and then having to retrain/refresh the airman when the task actually enters into the task inventory would be much higher than waiting to train the task as it begins to emerge. Also, career fields in which tasks emerge and perish very rapidly may be trained more efficiently through a combination of brief formal schooling and more extensive OJT. Because airmen in these career fields have tasks inventories that change quickly, OJT (whether structured or not) is probably already occurring at a high level. Time and money spent in training these tasks during the formal school could be ill spent.

Finally, a strength of this type of analysis is that it would provide information on a monthly continuum. An interesting application of survival analysis would be to link task emergence and task perishment to provide more information on when and by whom a task, or group of tasks, is performed in a career field.

REFERENCES

Christal, R.E., & Weissmuller, J.J. (1988). Job-task inventory analysis. In S. Gael (Ed.), *The job analysis handbook for business, industry, and government*, (Vol II), (pp. 1036-1050). New York: Wiley.

Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, Series B, *34*, 187-220.

Cox, D.R., & Oakes, D. (1985). *Analysis of survival data*. New York: Chapman & Hall.

Darden, R.W., Hampton, D.R., & Boatwright, E.W. (1987). Investigating retail employee turnover: An application of survival analysis. *Journal of Retailing*, *63*, 69-88.

Goldstein, R., Andersson J., Ash A., Craig, B., Harrington, D., & Pagano, M. (1989). Survival analysis software on MS/PC-DOS computers. *Journal of Applied Econometrics*, *4*, 393-414.

Kaplan, E.L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *American Statistical Association Journal*, *53*, 457-481.

Kiefer, N. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, *26*, 646-679.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, *35*, 139-156.

Lawless, J. (1982). *Statistical models and methods for lifetime data*. New York: Wiley.

Levine, E.L., Ash, R.A., Hall, H., & Sistrunk, F. (1983). Evaluation of job analysis methods by experienced job analysts. *Academy of Management Journal*, *26*, 339-348.

Morita, J.G., Lee, T.W., & Mowday, R.T. (1989). Introducing survival analysis to organizational researchers: A selected application to turnover research. *Journal of Applied Psychology*, *74*, 280-292.

Oakes, D., & Dasu, T. (1990). A note on residual life. *Biometrika*, *77*, 409-410.

Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies*, *57* 167-182.

Tuma, N., Hannan, M., & Groeneveld, L. (1979). Dynamic analysis of event histories. *American Journal of Sociology*, *84*, 820-854.

APPENDIX: SAS COMMANDS

```
title 'Simulated Study of 1000 Airmen';
*----------------------------------------------------------*
| This data step reads in the information on deaths and    |
|   censors.  Censored data points are marked with a       |
|   negative sign                                          |
*----------------------------------------------------------*;
Data perish;
        filename in 'work.dat;1';
        infile in;
        input x @@;
                if x > 0 then x = x + 12;
                else x = x - 12;
        y = x;
        censor = (x<0).
        x=abs(x).
        cards;
;
Proc lifetest data=perish outs=test plots=(s,ls) ;
        time x*censor(1)
;
*----------------------------------------------------------*
| Proc freq is used to count the number of deaths/         |
| censors per month.                                       |
*----------------------------------------------------------*;
Proc freq data=perish;
        tables y /out=j noprint;
*----------------------------------------------------------*
| This data step converts the censored data points from    |
| negatives to the month plus .5.  This will allow all     |
| of the censored data points to be subtracted from the    |
| risk set after the current month.                        |
*----------------------------------------------------------*;
Data haz1;
        set j;
        z=1.
        x=y;
        if x<0 then x=abs(x)+.5;
        keep z x count;
Proc sort data=haz1 out=k;
        by x;
*----------------------------------------------------------*
| This procedure is used to find the initial risk set.     |
*----------------------------------------------------------*;
Proc means data=j;
        var count;
        output out=l sum=s;
Data z;
        set l;
        z=1.
*----------------------------------------------------------*
| This data step calculates the risk set and the hazard    |
| value for each month.                                    |
*----------------------------------------------------------*;
Data haz2;
        merge z k;
        by z;
        if _n_=1 then do,
                left=0; r=s; end;
        left=left+count;
        r=s-left;
        risk=r+count;
        if x=round(x) then do;
                hazard=count/risk; end;
        else delete;
        retain left;
        keep x hazard;
*----------------------------------------------------------*
| This data step reduces the survival estimates so that    |
| there is only one value per month.                       |
*----------------------------------------------------------*;
Data mlrf1;
        set test;
        surv=survival;
        if _N_=1 then cut=0;
        if _N_>1 then do;
        if chop=survival then delete;
        else chop=survival;
        retain chop;
        end;
*----------------------------------------------------------*
| Proc means is used to find the sum of the survival       |
| estimates.                                               |
*----------------------------------------------------------*;
Proc means data=mlrf1 sum;
        var survival;
        output out=stats sum=s;
*----------------------------------------------------------*
| This data step sums the survival estimate from time 0    |
| to the current month.                                    |
*----------------------------------------------------------*;
Data mlrf2;
        if _N_=1 then stot=0;
        set mlrf1;
        stot=stot+survival;
        retain stot;
        a=1;
Data a;
        set stats;
        a=1;
*----------------------------------------------------------*
| This data step subtracts the sum from time 0 to the      |
| current month from the sum of all the survival           |
| estimates, thus computing the sum from the current       |
| month to the last month.                                 |
*----------------------------------------------------------*
Data mlrf3;
        merge a mlrf2;
        by a;
        If survival=0 then meanlife=0;
        else meanlife=(s-stot)/survival;
        keep x survival meanlife;
Data print;
        merge mlrf3 haz2;
        by x;
Proc print data=print;
```